

Predicting League of Legends Match Results Based on Machine Learning

DongHua Wang

January 3, 2024

League of Legends (LoL) is a highly popular multiplayer online competitive game, featuring intricate game mechanics and team cooperation, making the prediction of match outcomes a challenging task. This study utilizes a dataset from Kaggle, comprising 9,879 ranked matches ranging from Diamond I to Master tier, to build a machine learning model predicting the ultimate winner, either the blue or red team, based on the features of the first 10 minutes of gameplay.

Through steps such as data loading, preprocessing, and feature engineering, we provided effective inputs for the model. For model selection, we opted for the Logistic Regression algorithm, achieving a model accuracy of 0.7277 through data splitting and training. This accuracy robustly supports predictions of the winning side, whether blue or red.

However, to further enhance model performance, we recommend exploring additional feature engineering methods, investigating alternative machine learning algorithms, and fine-tuning hyperparameters. The introduction of deep learning models is also a promising avenue to better capture the complex relationships within the game. Through these improvements, we anticipate increasing the model's predictive accuracy for future matches, offering valuable insights for game development and enhancement.

Keywords: League of Legends, Machine Learning, Match Outcome Prediction, Feature Engineering, Logistic Regression, Accuracy

1 Introduction

League of Legends (LoL), as a globally popular multiplayer online competitive game, has attracted millions of players worldwide. Its unique game mechanics, diverse hero selection, and strategic team cooperation make each match an unpredictable and challenging battle. For the esports community, accurate predictions of match outcomes not only enhance the gaming experience but also offer valuable insights for adjustments in professional teams and game balance.

This study focuses on the first 10 minutes of League of Legends matches, a crucial period often determining the direction of victory or defeat. Through detailed analysis of data from 9,879 matches in high-ranking tiers (Diamond I to Master), we aim to build a machine learning model capable of predicting the final result based on the game features of the first 10 minutes. This predictive model holds the potential to be a beneficial tool in the esports domain, providing players with deeper game insights and offering opportunities for professional players and teams to enhance tactics and strategies.

In our exploration, we conducted data loading and preprocessing, selecting the logistic regression algorithm as the initial modeling tool. Through validation on training and testing sets, we achieved a model accuracy of 0.7272. However, we acknowledge that this is just a starting point. Future efforts should include more complex feature engineering, in-depth exploration of different algorithms, and a comprehensive evaluation of model performance.

This study not only offers a new perspective on predicting League of Legends match results but also opens up new possibilities in applying machine learning to the field of esports. Through this work, we aspire to make meaningful contributions to the League of Legends community, esports research, and the application of machine learning in the gaming industry. Source code is shared in a Github repository. <https://github.com/wdh7022/Predicting-League-of-Legends/tree/main>

2 Related Work

2.1 Domestic Research Status

In China, with the rapid development of data science and machine learning, research in the field of esports has gradually gained attention. League of Legends (LoL), as a widely popular multiplayer online competitive game, has become a focal point of research. Researchers delve into the analysis of match data, exploring key factors such as player behavior, team strategies, and match outcomes. Additionally, some studies attempt to use machine learning techniques to predict the outcomes of esports matches. By modeling and analyzing pre-match data, researchers seek to identify crucial features influencing match results, thereby providing more accurate predictions. There is also research focused on the balance of in-game elements, offering guidance to game developers for improving the gaming experience.

Internationally, there is a substantial interest in research within the field of esports, particularly with games like League of Legends taking the forefront. The application of machine learning in esports is even more widespread, with researchers attempting to predict match outcomes, analyze individual skills, and study team coordination. Furthermore, foreign researchers explore deep learning technologies in esports, such as neural networks, to better capture the intricate relationships within the game. Data visualization has also emerged as a research hotspot, presenting key match data through charts and graphics to make match analyses more intuitive and understandable. Overall, research efforts both domestically and internationally strive to provide a deeper understanding and advanced analytical methods for the field of esports.

2.2 Deep Learning

In the realm of competitive gaming, particularly in League of Legends, researchers have delved into the application of machine learning to forecast match outcomes. This endeavor involves employing various methods to enhance the accuracy of predicting results and gain deeper insights into the underlying patterns of the matches. Here are six prominent approaches within this domain:

2.2.1 Logistic Regression:

Logistic regression, serving as a linear model, proves invaluable in addressing binary classification problems, a common scenario in the field of esports, particularly in games like League of Legends. In this context, scientists harness the power of logistic regression to establish predictive models that correlate match outcomes with a plethora of features.

League of Legends, being a highly complex multiplayer online battle arena game, presents researchers with the challenge of predicting match results based on various in-game factors. Logistic regression, with its ability to handle binary outcomes effectively, becomes a prominent tool in this endeavor. By leveraging logistic regression models, scientists aim to uncover relationships between diverse features, such as player statistics, team composition, and early-game events, and the ultimate outcome of a match.

The logistic regression framework allows researchers to quantify the impact of each feature on the likelihood of a specific outcome, whether it be a victory for the blue or red team. This statistical approach not only provides a predictive model but also offers insights into the relative importance of different factors in determining match results.

Furthermore, logistic regression serves as a foundation for exploring more advanced machine learning techniques. Researchers often extend their investigations to incorporate additional algorithms, feature engineering methods, and hyperparameter tuning, aiming for continuous refinement and improved predictive accuracy. In essence, logistic regression acts as a crucial building block in the broader landscape of predictive modeling within the dynamic realm of esports research, contributing to a deeper understanding of the factors influencing game outcomes in League of Legends.

2.2.2 Decision Tree Algorithm:

Characterized by a tree-like structure representing decision processes, decision trees prove to be instrumental tools in modeling decision-making, particularly in the context of multiplayer online games

like League of Legends. The inherent structure of decision trees lends itself to intuitiveness and interpretability, making them valuable for analyzing and predicting outcomes in matches.

In a decision tree, each node corresponds to an attribute, and each branch signifies a decision based on that attribute. This hierarchical representation allows for a clear and logical breakdown of decision-making processes. In the specific case of League of Legends research, decision trees are employed to analyze key features of matches.

Researchers leverage decision trees to explore and understand the relationships between various in-game features and the ultimate outcome of a match. Attributes such as player statistics, team composition, and early-game events are represented in the decision tree structure. By traversing the tree, one can interpret the decisions made at each node, gaining insights into the factors that contribute most significantly to match results.

The appeal of decision trees in League of Legends analysis lies in their ability to provide a transparent and interpretable framework. Players, analysts, and researchers can easily comprehend the decision paths and identify the critical features influencing match outcomes. This transparency makes decision trees not only a powerful modeling tool but also a valuable asset for enhancing the understanding of the intricate dynamics at play in the game.

2.2.3 Support Vector Machine (SVM):

Support Vector Machines (SVM), known for their robustness in classification and regression analyses, prove to be particularly effective when dealing with non-linear relationships. In the realm of League of Legends research, SVM emerges as a powerful algorithm applied to model match data, providing more accurate classification of match results.

SVM operates by finding the optimal hyperplane that maximally separates data points belonging to different classes. This capability to handle non-linear relationships makes SVM well-suited for capturing the complex interactions and dependencies present in multiplayer online games like League of Legends.

In the context of match outcome prediction, SVM analyzes various features such as player performance metrics, team composition, and early-game events. By mapping these features into a high-dimensional space, SVM strives to create a hyperplane that distinctly classifies matches into different outcomes, whether it be a victory for the blue or red team.

The strength of SVM lies in its ability to navigate intricate relationships within the data, allowing for more accurate predictions even when faced with non-linear and complex patterns. This makes SVM a valuable tool in the arsenal of techniques employed by researchers and analysts striving to enhance the predictive modeling capabilities in the dynamic landscape of esports research.

2.2.4 Deep Learning and Neural Networks:

Deep learning techniques, characterized by the construction of deep neural networks, play a pivotal role in comprehending intricate non-linear relationships. In the dynamic context of League of Legends, researchers harness the power of neural networks for predicting match outcomes, aiming to capture the complex features of the game more effectively.

Neural networks, particularly deep ones, consist of multiple layers of interconnected nodes that process and transform input data. This architecture enables the network to automatically learn intricate patterns and representations from the data, making it well-suited for capturing the nuanced relationships present in multiplayer online games.

In the specific application to League of Legends research, neural networks are employed to analyze diverse in-game features such as player statistics, team composition, and early-game events. By training on historical match data, the neural network learns to recognize complex patterns and dependencies, ultimately enhancing its ability to predict match outcomes.

The adaptability of deep learning models allows them to uncover hidden structures within the data, providing a more nuanced understanding of the factors influencing match results. The ability to automatically extract relevant features and discern non-linear relationships positions deep learning as a valuable technique in the pursuit of advancing predictive modeling capabilities in esports research.

As researchers continue to explore the application of deep learning in the gaming domain, these techniques hold the potential to unravel deeper insights into the intricate dynamics of League of Legends matches, contributing to a more comprehensive understanding of the game's competitive landscape.

2.2.5 Ensemble Learning:

Ensemble learning methods, exemplified by techniques like random forests and gradient boosting trees, serve as powerful tools that amalgamate multiple models to enhance overall performance. In the domain of League of Legends research, these ensemble methods find application by integrating various models, adapting to the complexity inherent in match data.

Random forests, a type of ensemble learning, involve constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This ensemble approach is effective in handling diverse features such as player statistics, team composition, and early-game events, providing a comprehensive analysis of match outcomes.

Gradient boosting trees, another ensemble technique, builds a series of decision trees sequentially, with each tree correcting the errors of the previous one. This iterative process allows the model to adapt and improve its predictive capabilities over time. In the context of League of Legends, gradient boosting trees prove valuable for capturing intricate patterns in match data and refining the predictive accuracy of outcomes.

Ensemble learning, by combining the strengths of multiple models, excels in scenarios where individual models may fall short. In the realm of esports research, where match data is rich and multifaceted, ensemble methods offer a holistic approach to predicting match outcomes. This adaptability and robustness make ensemble learning an essential component in the toolkit of researchers and analysts striving to unravel the complexities of League of Legends matches.

2.2.6 Clustering Algorithms:

Clustering algorithms, exemplified by K-means clustering, play a crucial role in the exploration of match data by grouping similar instances together and revealing patterns within the dataset. In the context of League of Legends research, clustering is employed to unveil similarities and differences among different types of matches, providing researchers with a nuanced understanding of match dynamics.

K-means clustering, a popular algorithm in this domain, divides the dataset into a predetermined number of clusters, with each cluster representing instances that share similar characteristics. In the context of League of Legends, this may involve grouping matches with similar player statistics, team compositions, or early-game events. By doing so, researchers can discern distinct patterns and trends that may not be immediately apparent through other analysis methods.

Through the application of clustering, researchers gain valuable insights into the diversity present in match data. It allows for the identification of distinct archetypes of matches based on various features, shedding light on the intricate dynamics of gameplay. This understanding is particularly beneficial for tailoring strategies, analyzing player behaviors, and enhancing overall comprehension of the multifaceted nature of League of Legends matches.

In summary, clustering algorithms serve as a valuable tool for organizing and categorizing match data, facilitating a deeper exploration of similarities and differences. By revealing patterns within the data, clustering contributes to a more comprehensive understanding of the diverse array of matches in the League of Legends esports landscape. The diversity of these methods underscores the widespread application of machine learning in the field of League of Legends. Researchers continually explore different approaches, pushing the boundaries of this field's development.

3 Experimental Setup

To construct an effective model for predicting League of Legends match outcomes, we implemented a series of experimental settings to ensure the model's performance and reliability. The following are key points of our experimental setup:

Data Loading and Cleaning: We obtained a dataset comprising 9879 matches from high-tier ranks (Diamond I to Master) from Kaggle. During data loading, we performed necessary cleaning, including handling missing values, addressing outliers, and standardizing features to ensure data quality and consistency.

Feature Selection and Engineering: We carefully selected features relevant to match outcomes, including but not limited to hero kills, deaths, gold, experience, and level. Through appropriate

Method	Quantity
DecisionTree	0.7019
support vector machines	0.7277
logistic regression	0.7272

Table 1: Result Accuracy.

preprocessing and engineering of features, our goal was to extract crucial information to assist the model in capturing match dynamics effectively.

Model Selection: Initially, we chose the logistic regression algorithm as our modeling tool, considering its applicability and interpretability in binary classification problems. Simultaneously, we explored various models such as decision trees, support vector machines (SVM), deep learning, etc., to investigate performance differences among different algorithms in predicting League of Legends match outcomes.

Training and Validation: We split the data into training and testing sets, with 80% used for model training and 20% for validation. To avoid overfitting, we employed techniques such as cross-validation for iterative model tuning and validation.

Performance Evaluation: Accuracy served as our primary performance metric, while also considering other metrics such as precision, recall, etc. Through a comprehensive evaluation of these metrics, we aimed to assess the model’s overall performance.

Through the aforementioned experimental setup, our objective was to establish a predictive model with generalization capabilities in real match scenarios, providing substantial contributions to the League of Legends community and esports research.

4 Experimental Result

In our study, we employed three distinct machine learning algorithms, namely Logistic Regression, Support Vector Machine (SVM), and Decision Tree, to predict League of Legends match outcomes. The accuracy rates for each algorithm are as follows: It is noteworthy that these algorithms exhibited similar overall prediction accuracy, hovering around 72%. However, this does not necessarily imply exceptional performance in predicting League of Legends match outcomes. The game’s complexity introduces numerous influencing factors, including individual skill levels, team cooperation, and strategic decisions.

While the predictive accuracy of these algorithms may not be ideal, we view this as an initial exploration. Future research directions may involve more sophisticated feature engineering, experimentation with alternative machine learning algorithms, and even the introduction of deep learning methods. By continually refining models and algorithms, we aim to enhance the accuracy of predicting League of Legends match outcomes, providing deeper insights for research and practical applications in the esports domain.

5 Conclusion

In conclusion, our study focused on employing machine learning models to predict League of Legends match outcomes based on data from the first 10 minutes of gameplay. We utilized three algorithms: Logistic Regression, Support Vector Machine (SVM), and Decision Tree. Despite achieving similar overall prediction accuracy, around 72

Our research has provided initial insights into the potential applications of machine learning in the esports domain. However, it underscores the need for further model refinement. Future efforts may concentrate on optimizing feature engineering, exploring advanced algorithms, and incorporating more complex techniques like deep learning. Through these endeavors, we aim to enhance the model’s ability to accurately predict League of Legends match outcomes, offering deeper insights for esports research and practical applications. As the esports industry continues to evolve, machine learning technologies hold promise as potent tools for understanding and optimizing match results.

6 Discussion

The final prediction accuracy of only 72% indicates the challenges faced by machine learning models in forecasting League of Legends match outcomes. This limitation primarily stems from the complexity of human behavior and the variability of the battlefield, making it difficult to fully capture and predict the gaming environment. Machine learning algorithms often encounter constraints when dealing with nonlinear, dynamic, and highly interactive gaming scenarios, constrained by model simplifications and specific feature limitations.

In the realm of esports, factors such as individual player skills, team collaboration, tactical decisions, and more contribute to a vast and intricate information network. The dynamic and highly subjective nature of these factors makes it challenging for models to accurately capture and quantify, thus reducing the reliability of predictions. Additionally, the randomness and uncertainty inherent in games pose challenges to accurate predictions, causing models to struggle when responding to unforeseen circumstances and unexpected changes.

Therefore, we acknowledge that machine learning models still have limitations in predicting human behavior, especially in highly dynamic and interactive gaming environments. While algorithms demonstrate excellence in handling large-scale data, they require more sophisticated modeling for complex human decision-making processes and tactical variations in games. Future research directions may explore more intricate model structures, comprehensive feature engineering, and a deeper understanding of in-game mechanisms to enhance the adaptability and accuracy of models. Throughout this process, it is crucial to maintain a humble recognition of the diversity of human behavior and the dynamism of gaming, realizing that achieving complete predictions of game results may always be a challenge.